AI upscaling Supporting Image Alignment in Photogrammetric Reconstruction

Alessandro Martinelli[1], Agnese Toni[1], Sara Brescia[1],

[1] Dept. Of Civil Engineering and Architecture, Pavia University, 27100 Pavia, Italy
alessandro.martinelli01@unipv.it, agne.toni@gmail.com, sara.brescia@unipv.it

Keywords: AI, Photogrammetry, Upscaling, Survey, Sharpening, Resolution.

Abstract

This research aims to investigate and analyze the contribution provided by preventive image processing using AI in the SfM data acquisition process. Specifically, the objective is to observe qualities and defects of "AI upscaling" integrated into the normal workflow of digital restitution, with the hypothesis that greater sharpness and resolution can lead to better alignments and better model generations. Other similar experiments have been carried out previously on the AI intervention in the photos to improve the alignments, but a generic procedure and with untrained public AI has not yet been experimented. In the first phase of the research, the most suitable AI upscaling method for the purpose is selected, processing example images and comparing them with the original and various settings of the AI parameters to maintain likeness. In the second phase, the procedure efficiency was evaluated in several ways: through the automatic evaluation of Agisoft Metashape's Image Quality, by directly comparing the resulting model with the standard digital reconstruction of the same objects, comparing with a model produced with the native AI upscaling of the return program, and finally comparing with a high-level survey without defects. This test was repeated for four case studies specifically diversified by several factors, including subject scale, level of detail, lighting, color, blur, and type of camera used. Clear improvements emerge in all cases of AI upscaling, even reaching successful reconstructions that would not have started with the canonical method. It can therefore be stated that this preventive upscaling procedure is decisive in cases of low quality or damaged datasets, while in cases of already qualitatively sufficient photos it can be a useful support for increasing the level of detail.

1. Introduction

This study proposes an innovative image restitution workflow within the photogrammetric survey process, hypothesizing that AI upscaling (resolution enhancement treatment) of photos can facilitate subsequent processing for the creation of point clouds and 3D models. To demonstrate this, the AI Automatic 1111 WebUI interface (or A1111) from Stable Diffusion (SD) will be used to upscale 4 sets of photos across different case studies, to support their processing into 3D models and evaluate their quality.

AI upscaling, besides increasing resolution, can indeed bring improvements such as sharpening but sometimes also adds non-existent details (albeit at an infinitesimal level, often negligible for architectural scale). Considering the goal of maximum plausibility and rigor in architectural surveying, these parameters have been carefully controlled, leading to the most realistic results possible.

As expressed in photogrammetric technology texts (Triggs et al., 2000), the image alignment process is based on the Structure from Motion (SfM) principle and the Bundle Adjustment (BA) method. According to this process, the photos are scanned to create a database of keypoints (or feature points) that will then be compared in the images to transform them into tie points, i.e. points common to the images and which can be inserted into three-dimensional space, generating the scattered cloud in case of sufficient stability. The recognition of these keypoints depends on their color and their relationship with the immediate surroundings, therefore sharpness and exposure are the fundamental parameters for defining the quality of the reconstruction processes, and in this research experiments were carried out to try to optimize these factors. By "sharpness" we mean a measure of localized contrast at the high detail level, which is minimal in out-of-focus images, and can be estimated automatically with special algorithms or evaluated empirically.

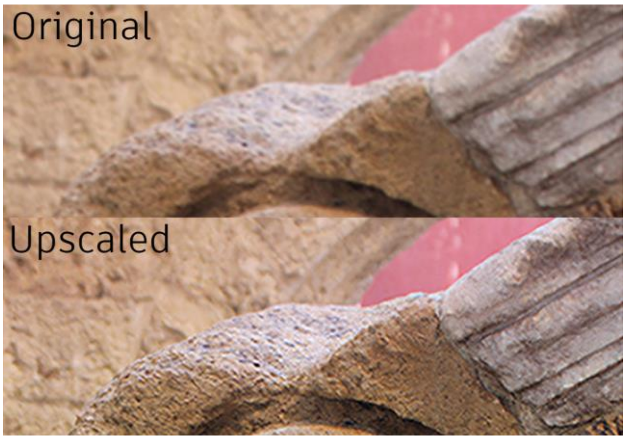

Figure 1. Comparison between an original photo and its upscaled version, where the sharpness is increased.

2. State of Art

2.1 AI in photogrammetry

Photogrammetry has played a crucial role in the analysis and digitization of architecture, cultural heritage, and design objects in recent decades. Indeed, numerous studies document field data acquisition for the construction of 3D models, which in turn serve as an important basis for subsequent interpretative phases and the development of reconstructive hypotheses (La Placa and Doria, 2022). Therefore, it is important to ensure their correct acquisition and optimize data processing (which this study addresses) so that cultural heritage can be transmitted effectively (Parrinello and Dell'Amico, 2019). With the advent of AI, several innovations have emerged in this field, particularly useful for achieving greater instrumental precision and automating data processing, up to the creation of entire datasets for the massive management of models (Doria et al.,

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-337-2024 | © Author(s) 2024. CC BY 4.0 License.

2021). Some of the most recent introductions in photogrammetry workflows have improved structure monitoring (Sabato et al., 2023), point cloud segmentation (Murtiyoso et al., 2022), or FEM and BIM implementation (Paduano et al., 2023). Other studies on AI in photogrammetry have enabled the reconstruction of the immaterial by compensating for the absence of image datasets with automatic interpretations (Condorelli, 2023). In this context, there appear to be few experiments on offline upscaling and sharpening tools to support photogrammetry, except for the case of Farella (Farella et al. 2022), who compared the sharpening capabilities between Photoshop and currently available online services, providing useful guidance for image enhancement. In other cases, to overcome sensor imperfections, dedicated AI has been directly trained for high-quality restitution in specific contents (such as space), yielding highly accurate results (Liesch et al. 2023). Recently, Pashaei (Pashaei et al., 2020) highlighted the capabilities of super-resolution on cases of downsampled images, compared with the originals; therefore in the 4 case studies adopted here, images with low resolution (LR) or defects from the start will be considered instead (without the intervention of downsampling), the enhanced version of which will be compared with the same surveys but carried out with high quality instrumentation and technique.

2.2 Case studies in Lombardia

2.2.1 Porta Nuova, of the city walls of Pavia: Porta Nuova (Fig. 02) is one of the smaller gates that flank the third city wall of Pavia, which is largely demolished today, with the 9 main gates serving as the most prominent evidence. Porta Nuova is associated with the "Spanish walls" erected between the 15th and 16th centuries, according to the plans of Giovanni Maria Olgiati to counteract the development of the artillery at the time, although the building itself dates back to the 11th century. In 2001 the building was restored by Prof. L. Jurina, with props and ribs that re-established its structural stability. The building features a simple quadrangular layout, with a wooden truss roof, resulting from various renovations in the modern era to cope with the sinking of the ground. The still bright red color of the masonry stands out from many other gates exposed to atmospheric agents over the centuries, offering a case study of greater visual interest. The two external facades facing northeast, including the arched entrance portal, two windows, and one of the monfores were chosen as the survey subject.

2.2.2 Certosa of Pavia: The Certosa is located in Pavia, close to Milan, in northern Italy, and is an important Carthusian monastery dating back to the 14th century. Its architecture combines Gothic and Renaissance elements, characterized by majestic towers and Baroque decorative details. The sculptural apparatus represents one of the major challenges in architectural survey, making it a useful comparison case. The case study focuses on the front facade of the building, with a symmetrical layout, with a triangular pediment surmounted by statues and sculptures. At the center of the portal stands a round arch, supported by fluted Corinthian columns, set in turn on bases decorated with geometric friezes and acanthus leaves, highlighting a Renaissance detail in their conception. Above the arch, there is a frame adorned with sculptures depicting biblical scenes and allegories, which add symbolic meaning to the portal. The sculptures are framed by a rich relief decoration, characterized by floral motifs, scrolls and masks, which emphasize the opulence and complexity of the entire façade.

Figure 2. The case studies: Porta Nuova, Certosa of Pavia, Ruins of S.M. del Buonconsiglio, Staircase spiral of San Domenico Maggiore church

2.3 Case studies in Puglia

2.3.1 Ruins of the church of *Santa Maria del Buonconsiglio:* The church of Santa Maria offers an exceptional chronological insight into one of Bari's most important periods, namely the transition from Byzantine to Norman rule. Its foundation is attributed by local historiography to as early as 755 AD (Andreassi and Radina, 1988). The Christian church is believed to have been constructed in two phases, the first in the 9th-10th centuries and the second in the 11th-12th centuries, as evidenced by a second layer of flooring. The second phase of the building's life may be related to its reconstruction following the damages inflicted by Guglielmo il Malo in 1156, which resulted in significant changes (Bertelli, 2004). In the 1990s, the structure, now in danger of collapse, was demolished, undergoing a significant restoration intervention in 1982. Today, the archaeological area remains unchanged from the arrangement defined during the restoration works of the 1980s, during which the remains of the church were recognized as of historical and archaeological interest.

2.3.2 Staircase spiral of the church of San Domenico Maggiore: San Domenico Maggiore, located in the northern area of the historic center of Taranto, was erected in the first half of the 14th century on the ruins of an archaic Greek temple. The exact date of its construction cannot be determined with certainty, nor is the promoter of its foundation known, despite an inscription on the facade guaranteeing its close connection with the noble family of Taurisano (Carannante, 2023). This main entrance elevation, which occurred towards the end of the 18th century to create the San Domenico slope, is filled by the Baroque staircase built in the center of the façade (Oliva, 2021). This staircase, bifurcated and with a double ramp, represents an interesting combination of perfectly geometric elements and others with a more irregular shape.

3. Methods

The experimentation was conducted on hardware with the following specifications: GPU GeForce RTX 3090 24GB, 128GB RAM, Intel i9 processor. The applied method is ideal for working with LR photos, allowing for upscaling to provide room for blur correction or sharpening without generating excessively large files (16K or more). An alternative to this process, if high-resolution photos are already available, is to use specialized techniques like "Add More Details - Detail Enhancer/Tweaker" LoRA, but this method is not guaranteed and will not be the focus of this study.

The experiments were divided into two phases. The first step of these was to select the most efficient method for upscaling realistic photos using SD AI maintaining maximum likelihood. The procedure, in addition to guaranteeing the maintenance of the contents, was carried out by avoiding third-party software or online services, adopting local computation with SD WebUI. In the second phase, the identified upscaling method was applied to the 4 case studies, comparing them with each other and evaluating their 3d model quality.

4. Experiments

4.1 Upscaling Method Selection

The first method analyzed is simply called "AI upscaling," which can be found in the "Extra" section of A1111. Regardless of the upscaling method used, however, the most basic factor to analyze and select is undoubtedly the upscaler model used. Today, there are hundreds of upscalers available for download online (1[1]), most of which are trained by amateurs. However, in order to obtain a rigorous and repeatable result, only those available natively in the WebUI were compared, along with the addition of an external one because it was considered one of the best in terms of sharpening: 4x-UltraSharp.

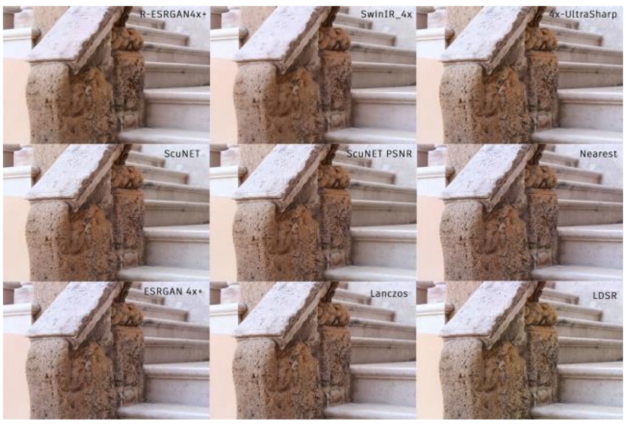

Figure 3. Processing of the same photo with different upscaler models.

From the comparison (Fig. 03), it emerges that the upscaler Lanczos' provides more detailed results than the others, except for LDSR which can produce significantly higher fidelity (both in detail and in color) but requires approximately 70 times the time of the other upscalers (12min-20sec compared to an average of 10.2sec). LDSR remains recommended in case of few photos and plenty of time available, but it is not applicable in these cases because it would turn one day of processing into three consecutive months. It is also interesting to note that LDSR can be applied both in x4 mode and in x2x2 mode (i.e., performing two upscalings at double the resolution instead of one at quadruple resolution), and the latter offers greater detail precision even though it tends to 'burn' the colors of the image (Fig. 04).

Figure 4. Comparison between x4 upscale (left) and LDSR x2x2 upscale (right), tested on LDSR.

In addition to Lanczos, however, the processing of "4x-UltraSharp" remains noteworthy because, compared to the former, it specifically maintains high sharpness and handles focal changes better, as shown in (Fig. 05). However, according to subsequent experiments, the level of detail offered by Lanczos allows for significantly better image alignments.

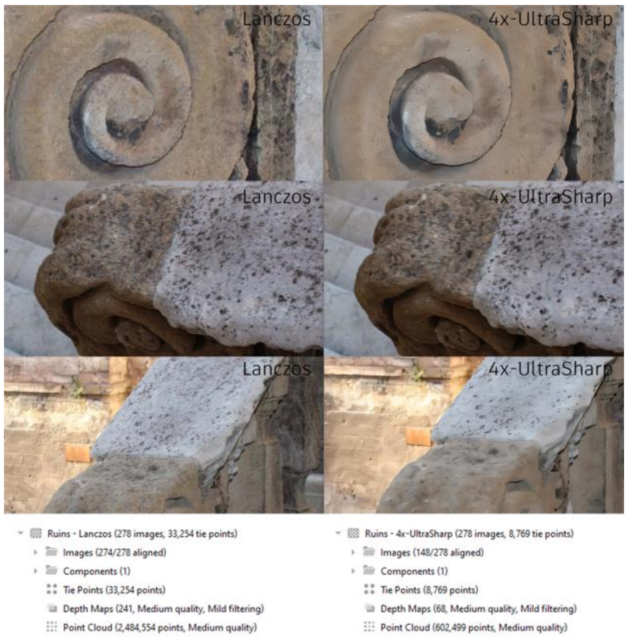

Figure 5. Comparison between the upscaler models "Lanczos" and "4x-UltraSharp." UltraSharp produces photos that appear sharper but with fewer elements and less depth.

Another essential comparison (Fig. 06) is between the checkpoint models applied to A1111, for which two SD models (1.5 and XL 1.0) and two amateur models relevant to the theme (Protogen V22 and Architecture Exterior) were chosen. Architecture Exterior and SD XL appear to be more faithful to color, while SD 1.5 and XL are more detailed, making XL the ideal choice.

[1] https://openmodeldb.info/

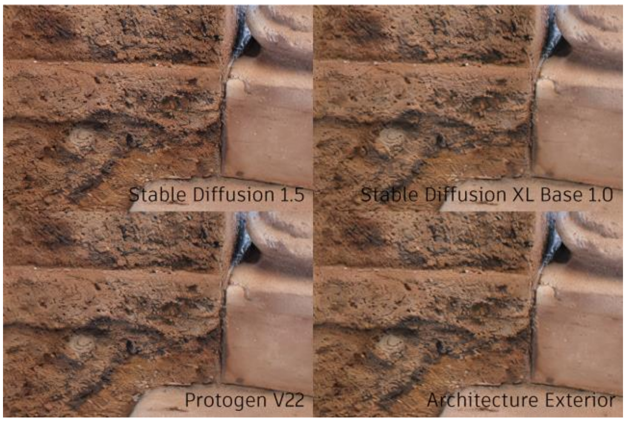

Figure 6. Comparison between the various Checkpoints applied in A1111. SD XL exhibits both a good level of detail and excellent color fidelity.

The second upscaling method used is the generative one called "SD upscale," which is a script that allows integrating the upscaling function into the "img2img" section of A1111. It also enables the generation of 'tiled' images, meaning one image tile at a time while still keeping the various parts 'seamless' (well integrated with each other), similar to how a bucket rendering would behave compared to a progressive rendering. This significantly expands the upscaling limits provided by img2img while still keeping the computational effort low. Currently, the script available natively is the "Ultimate SD Upscaler," and adopting it adds a considerable amount of parameters to analyze. The first of these is the "sampling method," which indicates the algorithm for denoising during image generation. The 4 sampling methods compared (Fig. 07) are the native ones considered most efficient: Euler, Euler A, DPM++2SAKarras, DPM++2MKarras. The sampler with the sharpest and most faithful results was found to be DPM++2SAKarras.

Figure 7. Comparison of the Sampling Methods. DPM++2SA Karras provides the best depth and sharpness.

Another fundamental parameter in the generation process is the "Denoise Strength" (Fig. 08), which determines how much noise margin to leave and consequently how much 'leeway' to give to the AI. Considering the objectives of this experiment, it is clearly necessary for this parameter to remain very low or zero. Therefore, all values between 0 and 0,3 were tested at intervals of 0,05. The results indicate that for the fidelity required, all values beyond 0,2 are not recommended, while results between 0 and 0,1 remain acceptable. Although the highest fidelity is achieved at 0, in cases of blurred photos, greater detail insertion can be useful.

Figure 8. Comparison between the original photo and different levels of denoise. The parameter ranges from 0 to 1, where 1 corresponds to 100% freedom given to the AI.

One last relevant factor is the choice to add the "ControlNet" function (Fig. 09), which performs a priori analysis of the image, extracting certain data chosen by the user, such as the depth map or object profiles. This function is usually used for specific purposes, and in this case, it provides minor improvements, albeit still present. Therefore, its use with the ‘tile’ pre-processor can be recommended, although it nearly doubles the processing time (2min-13sec compared to 1min-13sec without using it).

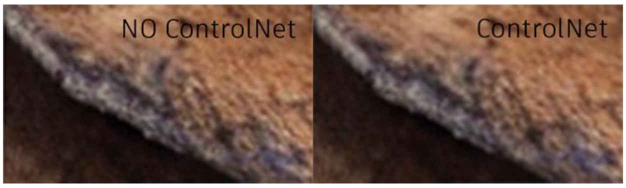

Figure 9. Comparison between SD upscaling performed without ControlNet and with ControlNet. The second image shows minimal differences in precision and color, which are nonetheless noticeable.

In conclusion, both automatic AI upscaling and SD upscaling produce extremely realistic results (Fig. 10), so close to reality that the differences between the products of the two methods are almost imperceptible. Ultimately, AI upscaling is preferred for slightly greater color fidelity and faster processing.

Figure 10. Final comparison between automatic AI upscaling and img2img with the SD Upscaler script and ControlNet activated. The differences are almost imperceptible, demonstrating the excellent performance of both methods.

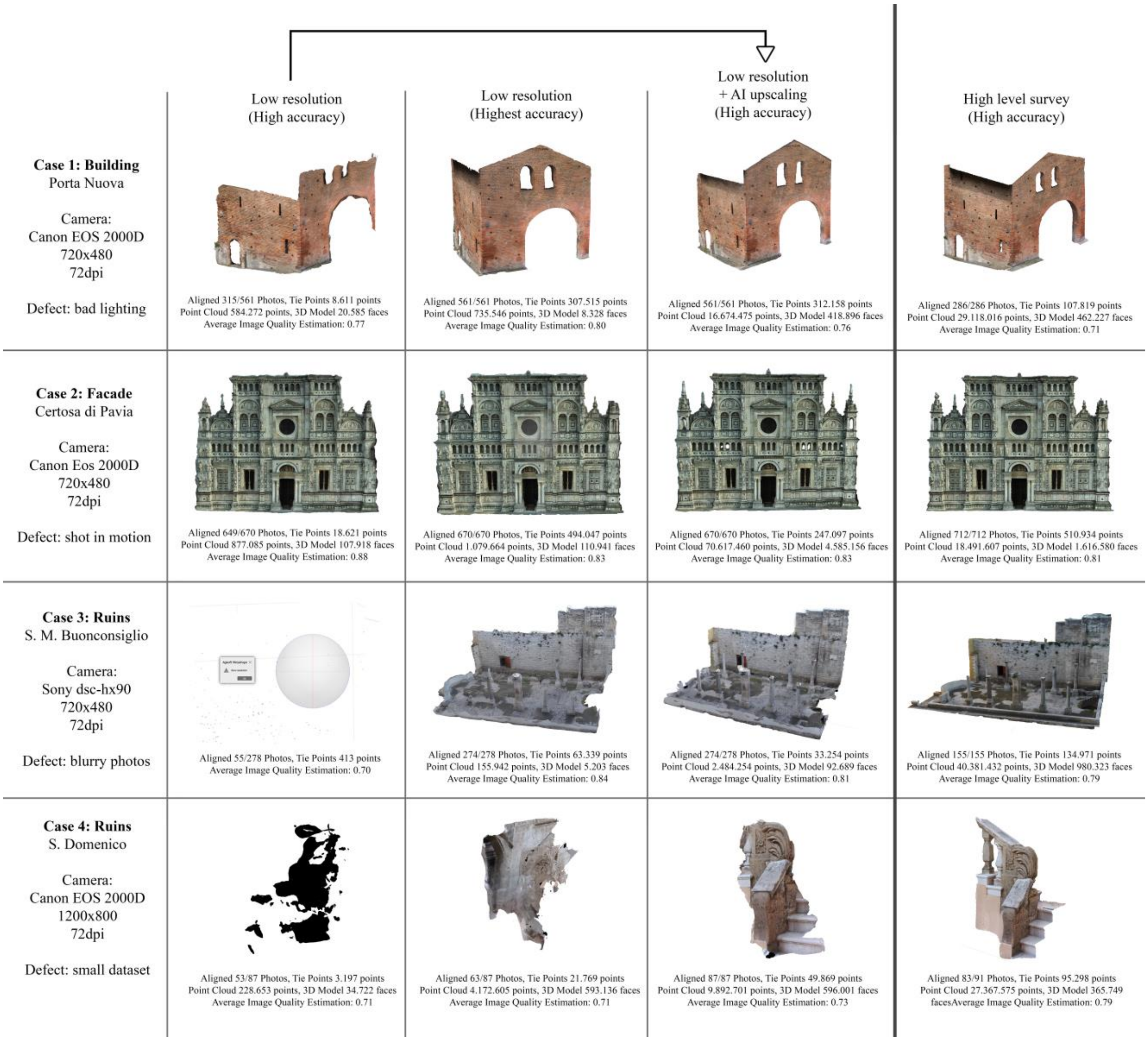

Figure 11. Comparison between the reconstructions of the original photos and the reconstructions of the upscaled/sharpened photos.

Figure 12. Comparison between model reconstructed with "Highest accuracy” alignment (i.e. native AI upscaling) (left) and model reconstructed with "High accuracy alignment" after AI upscaling (right). Slight distortions can be seen in the circled areas, which are however absent in the models generated with the proposed technique.

.

4.2 Photogrammetric processing

Once the best upscaling procedure was selected for the objectives of this research, four photographic sets were carried out on different case studies, which differ in various aspects (subject scale, location, level of detail, lighting, color, type of camera used, number of shots and intent of the photogrammetric campaign) and exhibit defects (blur, bad lighting, different levels of LR) that make the reconstruction difficult or impossible, thus lending themselves to improvements.

The differentiation of case studies based on scale and objective is a fundamental factor, because it defines what is evaluated or neglected in the final observation of the 3D model: for example, the color variety will be less important in a structural digital twin of a building, rather than in the orthomosaic of a valuable painting, and for this reason they were considered cases of both archaeological and architectural interest.

All photos were upscaled using the “AI upscaling” method (SD XL checkpoint and Lanczos upscaler) in 'batch' mode (i.e. one folder at a time, applying the effect to all the files contained) making the proposed process very quick, with an average of 6.5 minutes per folder. The process increased the resolution of the images by 4 times, with 0 denoise, thus without adding unnecessary details. The resulting images were brought into Metashape and processed with high accuracy both for alignment and for point cloud and model construction.

All 4 models were generated without errors, and were then compared (Fig.11-12) with:
- the model generated from the same photos without AI upscaling (column 1)
- the model generated from the same photos without AI upscaling but with the "Highest accuracy" alignment method (instead of "High accuracy") since this also uses a preventive upscaling treatment (column 2).
- the model generated from a different set of photos of the same subject, performed with high quality tools (Sony dsc-hx90 camera at 6000x4000 72dpi resolution) and techniques, at a professional level, keeping approximately the same number of shots (column 4).

5. Results

The processing of the original shots (in LR) showed failed, partial, or defective alignments, while the upscaled datasets had an average alignment percentage of over 99%, generating plausible and understandable models. It can be noted that the AI intervention always allows more photos to be aligned compared to the normal workflow, and furthermore it stands out for maintaining high resolution even in the generation of point cloud and 3D model.

In order to systematically evaluate the quality of the original and upscaled photos an original Metashape tool, "Estimate Image Quality", was also used, which offers a value from 0 to 1 for each individual image in the dataset, of which it was then averaged for each reconstruction carried out. According to the user manual, "the (image quality) value of the parameter is calculated based on the sharpness level of the most focused part of the picture[2]". The results show that each case study settles on stable, similar and always more than sufficient values (around 0.8). Interestingly, the cases of AI upscaling do not stand out in any case, and indeed the algorithm evaluates them similarly to LR images, although the final models present fewer defects than the others. No dataset, however, obtained ratings below the treshold of 0.5 which the manual identifies as not recommended.

When directly comparing the final 3D models, however, the following were taken into consideration: completeness of the product, consistency with the original, quantity of aligned cameras, number of points/polygons of the 3D model, quantity of noise, correct colouring.

The comparison with the "Highest accuracy" version, however, is more subtle since the preventive AI treatment carried out by Metashape allows an effective alignment like the one proposed, but generates a point cloud that lends itself less well to the generation of the 3D model, as it is possible to see from the comparison in Fig. 12: some parts, especially at the ends of the volumes and the edges, are distorted or missing.

The comparison with the high quality architectural survey, on the other hand, highlights imperfections and room for improvement also in the proposed technique, both in terms of quantity of shadow areas and resolution of the model. In fact, we can see that in some cases, such as that of Porta Nuova, fewer photos taken with the right technique were enough to obtain a better quality of the model than all the other post-processed cases.

However, the AI upscaling technique was found to be the most efficient of post-production interventions, useful in cases where the return to the campaign phase is no longer possible, and therefore only low quality photos are available.

6. Conclusions

The results of the research demonstrate the effectiveness of the proposed new method, with high potential and likely resolution for in all digital reconstruction processes in which the divergence from reality must not be less than 2-3 mm. The variety of case studies and the use of an AI that is not specifically trained show how this technique can be applied not only to cases of low image quality but also to cases of low sharpness, poor alignment, unsuitable equipment, or adverse environmental conditions. Furthermore, the introduction of this technique into normal workflows takes only a few minutes with sufficiently powerful hardware, making it easy and optimal in all contexts. This speed of use and the free accessibility of the necessary tools largely open the field of possible applications, from the architecture tested here to archaeology or the entertainment industry. The potential of this research leaves ample space for subsequent in-depth research, such as an analysis of the change in the tie points of individual images or their intrinsic parameters in camera calibration, in the comparison between LR and upscaled alignments. In the field of AI upscaling, however, possible expansions of the experiments are suggested such as comparison with dedicated AI training, or the use of LoRAs, or experimentation with image enhancement in other survey contexts.

Acknowledgements

All photos shown are taken by the authors for this research, except the ones of the Certosa di Pavia Facade, acquired in previous survey campaigns (2017-2020) by DAda-LAB research team. All the AI re-processing was performed by the authors with offline models.

[2] https://www.agisoft.com/downloads/user-manuals/

References

Andreassi G. & Radina F., 1988. Archeologia di una città. Bari dalle origini al X secolo. Catalogo della mostra tenuta a Bari – Santa Scolastica, 1988, Bari.

Bertelli G., 2004. Puglia preromanica: dal V secondo agli inizi del'XI, Milano: Jaca Book

Carannante A., 2023. *Note per lo studio dell'architettura delle prime fondazioni mendicanti in area pugliese (XIII-XIV secolo)*, in Colaceci S., Ragione R., Ravesi R., Rappresentazione, Architettura e Storia. La diffusione degli ordini religiosi in Italia e nei Paesi del Mediterraneo tra Medioevo ed Età Moderna, Atti del Convegno Internazionale, 10-11 maggio 2021, Roma.

Condorelli, F., 2023. *Image retrieval for 3D modelling of architecture using AI and photogrammetry*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 441-446.

Doria, E., Carcano, L., & Parrinello, S. (2021). *Object Detection Techniques Applied to UAV Photogrammetric Survey*. GiordanoA. RussoM. SpalloneR. (Eds.), Representation Challenges: New Frontiers of AR and AI Research for Cultural Heritage and Innovative Design. FrancoAngeli, 10.

Farella, E. M., Morelli, L., Rigon, S., Grilli, E., & Remondino, F., 2022. *Analysing key steps of the photogrammetric pipeline for Museum artefacts 3D digitisation.* Sustainability, 14(9), 5740. doi.org/10.3390/su14095740

La Placa, S., & Doria, E. (2022). *Reliability of DTMs obtained with mobile fast surveys techniques.* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 46, 299-306.

Liesch, A., Suwinski, P., Liu, B., Chernykh, V., & Janschek, K., 2023. *AI-Based LiDAR/Camera Data Fusion to Enable High-Resolution 3D Surface Reconstruction for Autonomous Asteroid Exploration Mission.* In AAS/AIAA Astrodynamics Specialist Conference.

Murtiyoso, A., Pellis, E., Grussenmeyer, P., Landes, T., & Masiero, A. (2022). *Towards semantic photogrammetry: generating semantically rich point clouds from architectural close-range photogrammetry.* Sensors, 22(3), 966.

Oliva L, 2021, Il complesso di San Pietro Imperiale, noto come S. Domenico, dall'età greca ai Predicatori: note per uno spaccato di identità urbana tra arte, civiltà e cultura materiale, in Derosa L., Triggiani M., Taranto la steel town dei beni culturali, Bari.

Paduano, I., Mileto, A., & Lofrano, E. (2023). *A Perspective on AI-based image analysis and utilization technologies in building engineering: Recent developments and new directions.* Buildings, 13(5), 1198.

Parrinello, S., & Dell'Amico, A., 2019. *Experience of documentation for the accessibility of widespread cultural heritage.* Heritage, 2(1), 1032-1044. doi.org/10.3390/heritage2010067

Pashaei, M., Starek, M. J., Kamangir, H., & Berryhill, J. (2020). *Deep learning-based single image super-resolution: an investigation for dense scene reconstruction with UAS photogrammetry*. Remote Sensing, 12(11), 1757.

Sabato, A., Dabetwar, S., Kulkarni, N. N., & Fortino, G. (2023). *Noncontact sensing techniques for AI-aided structural health monitoring: a systematic review*. IEEE Sensors Journal, 23(5), 4672-4684.

Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W., 2000, *Bundle adjustment—a modern synthesis.* In Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings (pp. 298-372). Springer Berlin Heidelberg.